

## Methylation of HPV and a tumor suppressor gene reveals anal cancer and precursor lesions

Attila T. Lorincz<sup>1</sup>, Mayura Nathan<sup>2</sup>, Caroline Reuter<sup>1</sup>, Rhian Warman<sup>1</sup>, Mohamed A. Thaha<sup>4,5</sup>, Michael Sheaff<sup>3</sup>, Natasa Vasiljevic<sup>1</sup>, Amar Ahmad<sup>1</sup>, Jack Cuzick<sup>1</sup> and Peter Sasieni<sup>1</sup>

<sup>1</sup>Centre for Cancer Prevention, Wolfson Institute of Preventive Medicine, Queen Mary University of London, EC1M 6BQ

<sup>2</sup>Homerton Anal Neoplasia Service, Homerton University Hospital NHS Foundation Trust, London E9 6SR

<sup>3</sup>Cellular Pathology, Barts Health NHS Trust, London E1 2ES

<sup>4</sup>National Bowel Research Centre, Blizard Institute, Queen Mary University of London, London E1 2AT

<sup>5</sup>Barts Anal Neoplasia Centre, Department of Colorectal Surgery, Surgery and Cancer CAG, The Royal London Hospital, Barts Health NHS Trust, Whitechapel, London E1 1BB

**Correspondence to:** Attila T. Lorincz, **email:** a.lorincz@qmul.ac.uk

**Keywords:** anal cancer, intraepithelial neoplasia, DNA methylation, high-risk human papillomavirus, HPV genotyping

**Received:** February 10, 2017

**Accepted:** May 06, 2017

**Published:** May 18, 2017

Copyright: Lorincz et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC-BY), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

### ABSTRACT

**We studied DNA methylation patterns of human papillomavirus (HPV) and tumor suppressor gene *EPB41L3* in 148 anal and perianal biopsies to determine whether high levels of methylation would be associated with anal intraepithelial neoplasia (AIN). The most prevalent HPV type was HPV16, detected in 54% of the 30 benign biopsies, 33% of the 43 low-grade AIN (lgAIN), 82% of the 59 high grade AIN (hgAIN) and 4 of the 5 anal cancers. A methylation score was developed ( $0.561 * HPV16me + 0.439 * EPB41L3$ ) which had increasing values with severity of disease: the mean was 8.1% in benign, 13.2% in lgAIN, 22.3% in hgAIN and 49.3% in cancers ( $p < 0.0001$ ). The methylation score as a triage classifier at a cut-off of 8.8 gave a sensitivity of 90.6% (95% CI: 82.8, 96.9), specificity of 50.7% (95% CI: 39.7, 61.6) and area under the curve of 0.82 (95% CI: 0.75–0.89) for separating hgAIN and cancer from benign and lgAIN biopsies. We conclude that methylation of HPV16 and *EPB41L3* show highly significant association with increasing severity of AIN and cancer and may be useful as biomarkers in anal disease.**

### INTRODUCTION

Human papillomavirus (HPV) infects a majority of people worldwide. Infection can occur at any age and can either be transient (usually resolving within a few years) or could be persistent and last for many decades [1]. High risk HPV (hrHPV) infection with types 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, and 68 in epithelial basal cells, especially in certain sites such as the uterine cervix, vulva, vagina, anus, and tonsils is an important risk factor for the development of squamous cell cancers and adenocarcinomas [1, 2]. Natural infections can produce immunity to identical and related HPV types while vaccination with virus-like-particles elicits a strong

humoral immune response that is an effective prophylaxis [3]. Persistence of hrHPV is a known strong risk factor for cervical cancer [4] and occurs in immunocompetent individuals but is more common in immunosuppressed patients, such as those infected by HIV [5]. The molecular mechanisms of transient versus persistent hrHPV infections have been only partially elucidated, but may involve differences in integration of the HPV genome into host DNA and DNA methylation [6, 7]. HPV DNA testing can identify almost all prevalent high-grade cervical intraepithelial neoplasia (CIN2 and CIN3) and cervical cancers in exfoliated cervical cells [8]. The test also has a good ability to predict incident disease several years in advance of clinical manifestation [9]. Recent widespread

recognition that hrHPV testing is much more sensitive than cytology has driven implementation of primary hrHPV screening for cervical disease in many countries [8].

Anal cancer has been growing in incidence in the past few decades, especially in women and also in men who have sex with men (MSM). Furthermore, anal cancer incidence is higher in HIV-positive MSM with approximately 100 cases compared to 25 cases per 100,000 in HIV-negative MSM [10, 11] and only 1.5 per 100,000 in men in general in the UK [12]. Most anal cancers have been associated with HPV16, while other hrHPV types such as HPV18, HPV31 and HPV33 seem to play a much smaller role in anal cancer than in cervical cancer [13]. High-grade AIN (alternatively called anal HSIL) is also associated with hrHPV, especially HPV16, and multiple HPV types are often reported in HIV-positive men [14].

Anal cytology sampling is problematic because anal folds may hide lesions and this has resulted in recommendations for more frequent sampling to compensate for poor sensitivity [15]. Normal, borderline or mildly dyskaryotic (also called ASCUS or LSIL) cytology is common in patients with hgAIN [16]. High-resolution anoscopy (HRA) is often used as the primary screening tool for high-risk populations in settings where resources can support such an intensive approach [17]. However, besides the high cost, using HRA to detect hgAIN has additional limitations such as a subjective result, availability of a trained HRA specialist and discomfort caused to the patient. Consequently, for decades, other methods of triage to biopsy and treatment have been actively sought to lessen the burden on the HRA clinics [18].

DNA methylation testing of HPV and human genes has been validated as an accurate method for detection of CIN2 and CIN3 [19–22]. Levels of methylation increase over time in women with persistent HPV16 infection and are maximal in patients with cancer [23, 24]. We investigated if a similar methylation test might be usefully applied to people with anal disease based on our *a priori* hypothesis that high levels of methylation at genomic positions associated with hgCIN would also be associated with hgAIN. Here, we focus on the methylation of host gene *EPB41L3* and the high risk viral types: HPV16, HPV18, HPV31 and HPV33. *EPB41L3* (Erythrocyte Membrane Protein Band 4.1 like 3) is a tumor suppressor gene that inhibits cell proliferation, promotes apoptosis and has been found to be highly methylated in many cancers such as lung, cervix, ovarian and breast [25–28].

## RESULTS

There were 30 biopsies with <AIN, 43 IgAIN, 59 hgAIN and 5 cancers among the anal samples and 11 biopsies of high-grade perianal lesions (Table 1).

## HPV genotyping

About a third (47/148) of samples were infected with multiple HPV types. 33% of anal biopsies with either <AIN, or IgAIN histopathology were not infected by hrHPV types while 53% and 33%, respectively, were HPV16 positive (Table 1). hgAIN anal and perianal biopsies were predominantly infected by HPV16 (83% and 91% respectively). A small proportion (9%) of anal and perianal hgAIN were infected with hrHPV types other than HPV16, i.e. 5 out of 59 of anal and 1 out of 11 of perianal biopsies. Also of note, all the perianal lesions were hrHPV positive. Only 9% of anal hgAIN biopsies were not infected by hrHPV (5/59). All cancers were hrHPV positive, four with HPV16 and one with HPV33.

## DNA methylation

The DNAm levels of *EPB41L3*, HPV16L1 and HPV16L2 were significantly different between the four groups and increased with severity of the lesions ( $p < 0.0001$ , Cuzick test for trend), but no significant differences in the likelihood ratios were found for methylation of HPV18, HPV31 and HPV33 and these latter markers were dropped from further analysis (Table 2). The univariable models (Table 2) investigating *EPB41L3*, HPV16L1 and L2 regions were all highly significant ( $p < 0.0001$ ). The bivariable logistic regression using *EPB41L3* and HPV16me was highly significant ( $p < 0.0001$ , Table 2) as was each variable on its own. The linearly combined DNAm score was derived from the bivariable model and calculated as follows:  $0.561 * \text{HPV16me} + 0.439 * \text{EPB41L3}$ . For all three variables (*EPB41L3*, HPV16me and the DNAm score), there was a highly significant trend of increased methylation with disease progression (Cuzick tests for trend,  $p < 0.0001$ ). Figure 1 shows the methylation of *EPB41L3*, HPV16me and the DNAm score.

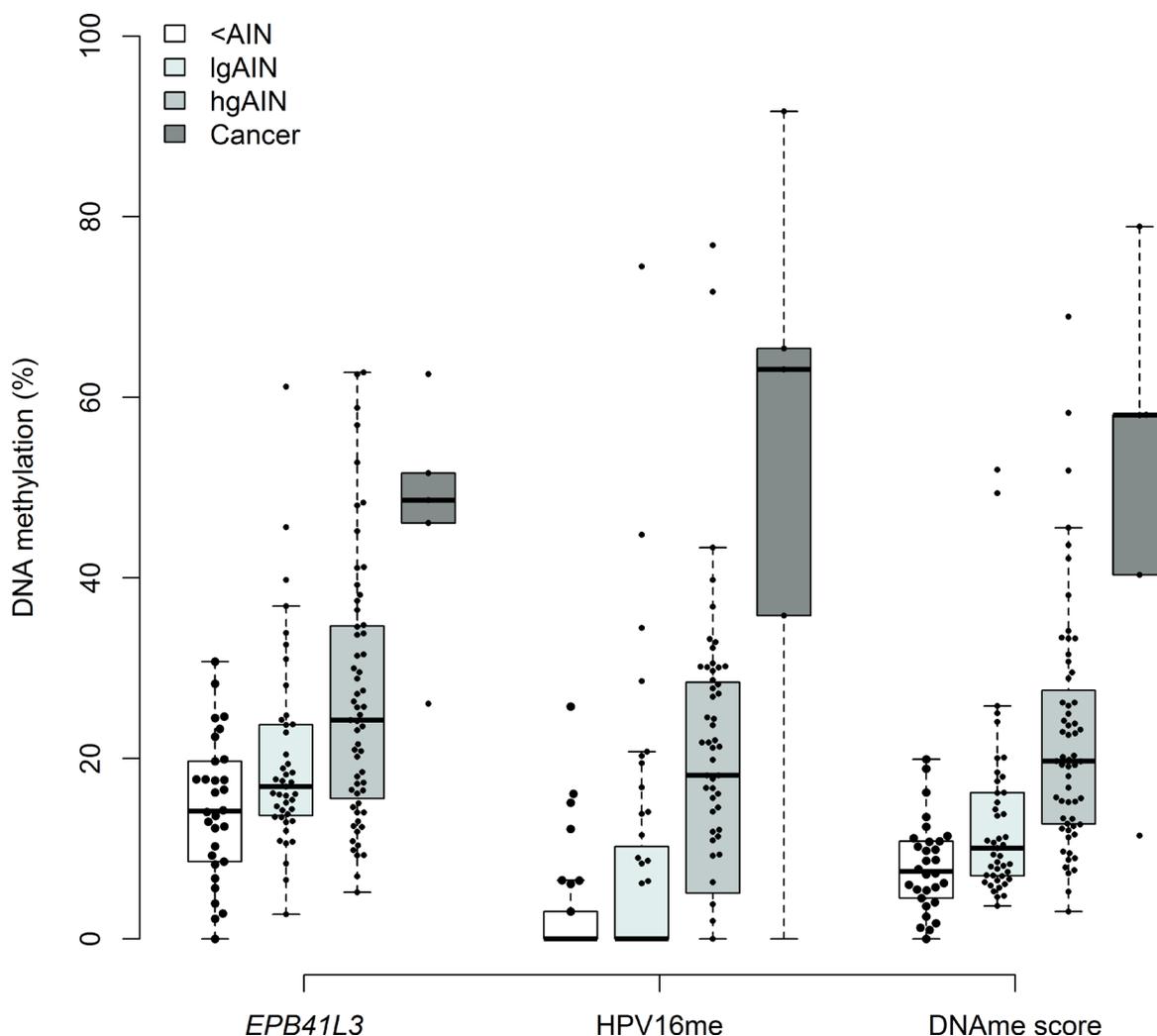
The ROC curves comparing the methylation levels in the <AIN and IgAIN samples to the hgAIN and cancer cases (Figure 2) had an AUC of 0.712 (95% CI: 0.624, 0.801,  $p < 0.0001$ ) for *EPB41L3*, 0.781 (95% CI: 0.705, 0.857,  $p < 0.0001$ ) for HPV16me and 0.821 (95% CI: 0.750, 0.892,  $p < 0.0001$ ) for the DNAm score. Figure 2 also shows the relative sensitivities and specificities of genotyping for HPV16 or genotyping for HPV16 and HPV18 combined. Supplementary Figure 1 shows ROC curves comparing missing DNAm values imputed by MICE versus single imputation. The chart indicates that there was no statistically significant difference in our interpretations of the data using either the multiple or single imputed data.

Sensitivity and specificity of the DNAm score to detect hgAIN and cancers and the proportion of positive samples identified by the DNAm score is shown in Table 3. At the 7.5 cut-off the DNAm score correctly

**Table 1: HPV typing data of anal and perianal samples by lesion type**

		HPV16	Other hrHPV types	lrHPV types or negative	Total
		<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)	<i>n</i>
Anal	<AIN	16 (53.3)	4 (13.3)	10 (33.3)	30
	lgAIN	14 (32.6)	15 (34.9)	14 (32.6)	43
	hgAIN	49 (83.0)	5 (8.5)	5 (8.5)	59
	Cancer	4 (80.0)	1 (20.0)	0 (0)	5
	Total	83 (60.6)	25 (18.2)	29 (21.2)	137
Perianal	hgAIN	10 (90.9)	1 (9.1)	0 (0)	11

Other hrHPV: high-risk and possibly high risk HPV other than type 16 (this includes HPV18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68); lrHPV: low-risk HPV (HPV6 and 11); AIN: anal intraepithelial neoplasia; lg: low-grade; hg: high-grade.



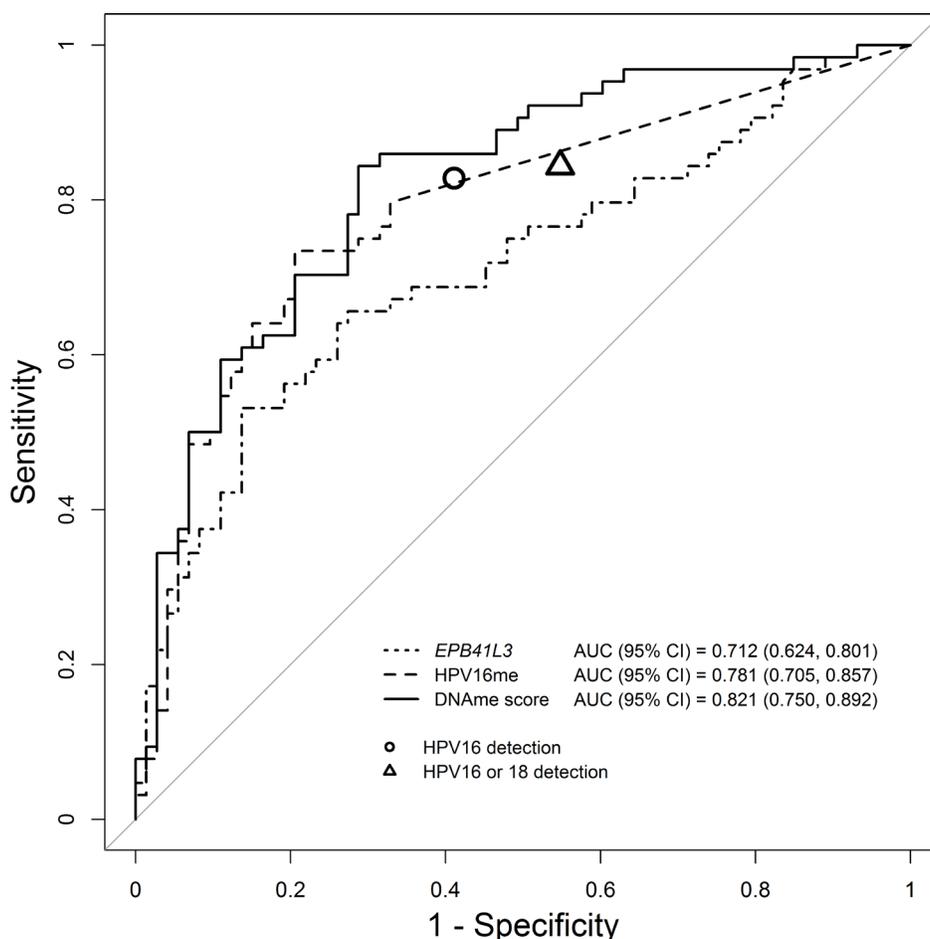
**Figure 1: Comparison of DNAm levels of *EPB41L1*, HPV16 and the DNAm score ( $0.561 \cdot \text{HPV16} + 0.439 \cdot \text{EPB41L3}$ ) in <AIN, lgAIN, hgAIN and cancer cases.** Perianal samples are not included in this figure. The top of box represents the upper quartile, bottom the lower quartile and line the median. The upper (lower) whisker extends to the largest (smallest) point that is not more than  $1.5 \times$  of the inter-quartile range from the upper (lower) quartile. All data points with a methylation value  $> 0$  are shown individually (black circle).

identified all the cancers and 95% of the hgAIN, while correctly classifying 33% of the lgAIN and 50% of the <AIN biopsies. In comparison, sensitivity of HPV16 genotyping to detect hgAIN and cancer was 83% (95% CI: 73–92) and specificity was 59% (95% CI: 48–70).

Supplementary Figure 2 shows a scatter plot of methylation of *EPB41L3* against HPV16me with the relationship having a weak Spearman correlation. We also investigated whether the DNAm score would be able to correctly identify the 11 high-grade perianal samples (Supplementary Figure 3) using cut-offs obtained with anal samples. Ten of the cases were correctly identified at the cut-off 7.5 and nine at the cut-off 8.8. Finally, we stratified the methylation data by HPV16 positivity and found that most of the predictive methylation information was in the HPV16 infected patients, which demonstrated that DNA methylation provided triage information in addition to the information given by HPV16 genotyping. In contrast little if any diagnostic contribution was seen for methylation in the HPV16 negative group (Supplementary Figure 4).

## DISCUSSION

The *a priori* hypothesis that high levels of methylation at genomic positions shown to be associated with hgCIN in our earlier research [19, 29] would also be associated with hgAIN has been confirmed, thus opening the way to methylation diagnostics of anal disease. Anal cytological abnormalities are poor predictors of hgAIN amongst HIV-positive patients [30]. Moreover, there is not a good correlation between cytology grades and histology grades. There are also substantial differences between pathologists in interpreting anal histology. In our study, we used a single pathologist (MS) who had extensive experience in anal pathology for all histology and the diagnoses were backed up by p16 staining when indicated [31, 32]. Goldstone et al. showed that 20% of patients with normal anal cytology and more than 30% with borderline cytology had hgAIN [33]. HPV DNA testing as a primary anal screen has the advantage of greater sensitivity than cytology and has an advantage of lower costs than screening with high-resolution anoscopy.



**Figure 2: Receiver operator characteristic and associated area under the curve (AUC) of DNAm of *EPB41L3*, HPV16 and the DNAm score.** The DNAm score (solid line) performed significantly better than *EPB41L3* (dotted line) or HPV16 methylation (dashed line) on their own. For comparison the circle represents the performance of HPV16 genotyping while the triangle represents the performance of combined genotyping for HPV16 or HPV18.

**Table 2: Univariable and bivariable logistic models of DNA methylation, age and gender**

Logistic models					
Univariable			Bivariable		
Markers	N	IqrOR <sup>a</sup> (95% CI)	Model LR $\chi^2$ (p)	OR <sup>a</sup> (95% CI)	Model LR $\chi^2$ (p) <sup>b</sup>
DNAme score <sup>c</sup>	137	5.454 (2.664, 11.154)	39.036 (4.16e-10)		
HPV16me	137	4.866 (2.454, 9.647)	31.573 (1.92e-08)	1.066 (1.030, 1.104)	31.573 (1.92e-08)
HPV16me L2	137	3.689 (1.914, 7.112)	25.466 (4.5e-07)		
<i>EPB41L3</i>	137	2.836 (1.721, 4.676)	22.292 (2.34e-06)	1.052 (1.012, 1.092)	7.463 (0.0063)
HPV16me L1	137	3.306 (1.708, 6.401)	14.286 (1.57e-04)		
Age	136	1.769 (1.069, 2.927)	5.221 (0.0223)		
HPV31me	137	0.101 (0.008, 1.348)	3.832 (0.0503)		
HPV33me	137	9.562 (0.478, 191.242)	1.531 (0.2160)		
Gender	136	0.576 (0.236, 1.407)	1.487 (0.2227)		
HPV18me	137	1.013 (0.982, 1.045)	0.692 (0.4054)		

<sup>a</sup>IqrOR and OR for a 1% change in DNA methylation, <sup>b</sup>variables added sequentially (first to last), <sup>c</sup>The score was developed as a linear predictor of the bivariable logistic regression, me: methylation.

DNA methylation is presented for the newly developed score (DNAme score), *EPB41L3*, HPV16, HPV18, HPV31 and HPV33. For HPV16, the methylation levels of two regions, L1 (HPV16me L1) and L2 (HPV16me L2), and the average of both regions (HPV16me), are presented. The DNAme score is calculated using  $(0.561*HPV16me+0.439*EPB41L3)$  to predict high-grade anal neoplasia. Odds ratios (OR), inter-quartile odds ratios (IqrOR), likelihood ratios (LR),  $\chi^2$  and *p*-values of the models are given.

**Table 3: Sensitivity and specificity at specified cut-offs to detect hgAIN and cancers and proportion of positive samples identified by the DNAme score**

Cut-off	% Sensitivity (95% CI)	% Specificity (95% CI)	$\chi^2$	<i>p</i> -value	Proportion of samples identified by the DNAme score			
					<AIN	lgAIN	hgAIN	Cancer
7.5	95.3 (90.6–100)	39.7 (28.8–50.7)	25.426	4.60e-07	0.50	0.67	0.95	1.00
8.8	90.6 (82.8–96.9)	50.7 (39.7–61.6)	27.872	1.30e-07	0.40	0.56	0.90	1.00
10.3	85.9 (76.6–93.8)	58.9 (47.9–69.9)	42.117	8.60e-11	0.20	0.40	0.85	1.00

The 95% confidence intervals (95% CI) were computed using 2000 stratified bootstrap replicates. AIN: anal intraepithelial neoplasia; lg: low-grade; hg: high-grade.

Another advantage of HPV screening is that anoscopy is a complex procedure with a need for extensive training of practitioners. However, the fundamental problem with HPV DNA testing that has precluded its widespread use in identifying anal precancer is poor specificity. For example, Salit et al. [17] showed that 88% of their HIV+ patients were HPV DNA positive for carcinogenic types. The specificity of HPV testing can be partially rectified by focusing on HPV16, which is one of the most common types found in anal cancer [34–37]. However, since HPV16 is not present in all anal cancers there is a

good chance that high-risk progressive lesions will not be detected. In our study, we found that most AIN were positive for hrHPV DNA and the majority of hgAIN were positive for HPV16, with relatively few positive for other hrHPV types. However, only 4 of the 5 anal cancers were positive for HPV16, the other being positive for HPV33, which shows the limitation of relying on HPV16 genotyping triage.

DNA methylation is a potential option that may offer greater improvements in triage specificity while retaining good sensitivity and importantly detect all the

cancers. In our study, all of the anal cancers were positive (i.e. above the cut-off) for DNA methylation, similar to what has been generally observed for cervical cancer. The DNAm score we developed is a multi-biomarker panel composed of three CpG sites within the *EPB41L3* gene and the late regions (L1 and L2) of HPV16. The AUC of the methylation score for separating <AIN and IgAIN from hgAIN and cancer was 0.82 (95% confidence interval: 0.75, 0.89,  $p < 0.0001$ , Table 3). Studies have shown that cervical cancers have higher levels of methylation than CIN3 [20]. All the anal cancers in our study were highly methylated and quite well separated from the hgAIN. This suggests the possibility that DNA methylation may be used to indicate AIN destined to progress to anal cancer from lesions that will regress or remain indolent [38–40].

It is notable that the triage information for hgAIN and cancer provided by the methylation score was significantly higher only in HPV16 positive people (Supplementary Figure 4); however, the cancer negative for HPV16 was strongly methylated (25%) for *EPB41L3*. These data demonstrate that DNA methylation provides significant diagnostic information for detecting hgAIN in addition to that provided by HPV16 genotyping and suggests that a methylation test may be used to detect essentially all the cancers. The methylation score in HPV16 negative people did not show a significant discriminating effect. However, we cannot be sure of the lack of value of methylation information in the HPV16 negative samples because the analysis was underpowered for this endpoint. These results indicate the need for larger studies and a search for more genes that could provide additional triage information in people infected with hrHPV types other than HPV16.

A limitation of the study is that we used FFPE biopsies. Our results need to be replicated in an adequately powered study of exfoliated anal cells because in routine practice hrHPV positive patients would have methylation tests performed on exfoliated cells collected by a swab or similar device. This would allow efficient triage to HRA, thus reducing costs, anxiety and possible over-treatment of low risk people. Our study had incomplete information on HIV status and the small size of the HIV-negative subset likely produced some ascertainment bias. Another limitation of our study is the use of patients (mostly MSM) recruited from two sites specializing in anal HPV-related disease in London. It remains to be seen if our results can be duplicated in other settings.

We conclude that high levels of DNA methylation are associated with hgAIN and anal cancer. This finding should be further explored to better understand the biological mechanisms and the value of DNA methylation testing as a molecular triage of hrHPV positive individuals for high-resolution anoscopy screening.

## MATERIALS AND METHODS

### Patients

A set of anal and perianal biopsy specimens were obtained from 148 patients (116 men, 31 women, and 1 person of unrecorded gender) of whom 94 were HIV positive, 40 were HIV negative and 14 had not been tested for their HIV status (Supplementary Table 1). The formalin-fixed paraffin-embedded (FFPE) biopsies were retrieved from the archives of the Homerton University Hospital and St Bartholomew's and the Royal London Hospital in London, UK, which are tertiary referral units where people with suspected hgAIN are referred for further management. Institutional approval (R&D number: GU1310) for the study was obtained prior to commencing any research work on the specimens.

In patients undergoing HRA, biopsies were obtained from areas of clinical interest that exhibited acetowhite changes (regardless of vascular changes) on 5% acetic acid application. Most of the biopsies showed morphological changes varying from slight to severe which distinguished them from the surrounding non-acetowhite normal epithelium. The main clinical endpoint for all comparisons was the histology result. The biopsies were graded histopathologically using the AIN terminology [41] which is in general use in the UK, acknowledging that recent recommendations for terminology in the US distinguish between low grade squamous intraepithelial lesion (LSIL equivalent to AIN1) and HSIL (equivalent to AIN2 and AIN3). The histopathological diagnoses were based on expert review by one of our team (MS) who has worked on anal neoplasia for more than 10 years. Diagnostic adjudication of difficult AIN cases was assisted by p16 staining of tissue sections [42] using the following rule: if the p16 result was positive (diffuse staining), the higher diagnosis (hgAIN) was assigned and if p16 was negative (focal, sporadic and negative staining), the lower diagnosis (IgAIN) was assigned. However, some of the biopsies showed only presence of HPV, and other biopsies showed slight changes that did not fulfil the criteria for IgAIN or hgAIN. These evidently non-normal biopsies are assumed as benign and were graded as <AIN.

### HPV genotyping

We used H&E sections annotated by MS as a guide to dissect areas of interest on the corresponding four unstained 5  $\mu$ m sections, as previously described [43]. If more than one lesion was present on a single section, we dissected the lesion with the highest grade. The dissected areas from the four sections were placed in the same tube. Dissected areas were deparaffinized using 80  $\mu$ L of hexadecane followed by a 5-minute incubation at 56°C. One hundred microliters of universal extraction buffer containing 50 mM Tris-HCl pH 8.0, 1 mM EDTA and

0.05% SDS [44] was added to tissues along with 10  $\mu$ l of Proteinase K (QIAGEN) and incubated overnight at 56°C followed by a one-hour incubation at 90°C. The lower phase was then transferred to a new tube and stored at -20°C before PCR.

The samples were tested using the PapType High Risk HPV Detection and Genotyping kit (PapType kit, Genera Biosystems Ltd) according to the manufacturer's instructions. The kit is able to detect 13 high-risk HPV types (HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, and 68) a possible high risk type (HPV66) and two low-risk types (HPV6 and HPV11). The PapType test was performed with 10  $\mu$ L of DNA in a final reaction volume of 20  $\mu$ L with the addition of 2% Tween 20. The PCR reaction amplifies a variable region of the L1 gene of the HPV genome. A fragment of the human cardiac myosin light chain gene (MLC-1) was co-amplified in the same reaction vessel as a quality and quantity control. For simplicity, we categorized the samples into three HPV genotype groups independently of whether they were singly or multiply infected: those infected by (1) HPV16, (2) by any other high-risk types (including HPV66) and (3) those HPV negative or infected only by a low-risk type (lrHPV). We did not combine HPV18 with HPV16 due to obvious differences in viral disease characteristics as noted in earlier studies [1, 45] and the fact that HPV18 does not seem to play such an important role in anal cancers [46].

### Methylation assays

Bisulfite conversions on 20  $\mu$ l of DNA extracts were done using EZ DNA Methylation Kit (Zymo Research, CA, USA) following the manufacturer's instructions. DNA methylation (DNAm) was measured by pyrosequencing for human biomarker *EPB41L3* and viral late genomic regions of HPV16 (CpG sites in the L1 region: 6367, 6389 and L2 region: 4238, 4247, 4259, 4268, 4275) and HPV18 (CpG sites in the L2 region: 4257, 4262, 4266, 4269, 4275, 4282), HPV31 (CpG sites in the L1 region: 6352, 6354) and HPV33 (CpG sites in the L2 region: 5557, 5560, 5566 and 5572) as previously described [47]. Amplification of CpG sites were carried out using PyroMark PCR kits (QIAGEN, Germany) with 4  $\mu$ l (8  $\mu$ l) of converted DNA for *EPB41L3* (HPV assays) in a 25  $\mu$ l volume with final concentration of reagents of 1 $\times$  for Coral Load and PyroMark mix, 0.2  $\mu$ M of PCR primers. PCR cycling conditions were 15 minutes at 94°C, followed by 45 cycles of 94°C, 54°C, 72°C each for 30 seconds and a final extension at 72°C for 10 minutes. The PCR products were pyrosequenced using a PyroMark™Q96 ID (Qiagen) instrument as previously described [48]. All pyrosequencing runs included a negative control and positive controls of known methylation level (0%, 50% and 100%) to allow standardized direct comparisons between different primer sets. For each marker, we calculated the average methylation level by taking the mean of all CpG positions.

### Statistical methods

Missing methylation values for HPV were imputed with the value of zero for any HPV negative sample. Missing methylation values for *EPB41L3* and HPV positive samples were imputed using a median regression with age as a predictor and DNA methylation as an outcome. All statistical analyses were performed on the imputed data set [49]. Out of the 137 anal samples, six had missing methylation values for *EPB41L3*, 22 for HPV16L1, 31 for HPV16L2, 5 for HPV18, 11 for HPV31 and none for HPV33. Of the 11 perianal samples, 4 had missing values for HPV16L1, 1 for HPV16L2 and 3 for HPV18. No value was missing for *EPB41L3*, HPV31 nor HPV33 (Supplementary Table 2).

Statistical analyses for the anal and perianal lesions were not combined. Spearman correlations were calculated between the markers. Since methylation values of HPV16L1 and HPV16L2 regions were correlated (Spearman  $r = 0.570$ ,  $p < 0.0001$ ), a variable called HPV16me was created by taking the geometric mean of DNAm levels of HPV16L1 and HPV16L2.

Univariable and bivariable logistic models were fitted for statistically significant genes with the outcome measures 0 = <AIN and lgAIN, and 1 = hgAIN and cancer. The likelihood ratio (LR)  $\chi^2$  statistic and its corresponding  $p$ -value, as well as the odds ratios (OR) with 95% confidence intervals were estimated and reported. Performance of the markers were assessed univariably by the LR  $\chi^2$  test and the area under the curve (AUC) with 95% confidence intervals (CI) using the Delong method [50]. Sensitivity and specificity were estimated at selected cut-offs (i.e. cut-off values giving the same sensitivity as HPV16 genotyping as well as at 90% and 95% sensitivity).

A combined DNAm score of *EPB41L3* and HPV16me was computed as a linear predictor of the fitted bivariable logistic regression. Confidence intervals for difference in sensitivities and specificities between the DNAm score at different cut-offs and HPV16 genotyping was computed using 2000 stratified bootstrap replicates as recommended by Carpenter and Bithell [51]. Cuzick tests for trend were applied to *EPB41L3*, HPV16me and the DNAm score to test for significant changes of methylation between the groups (<AIN, lgAIN, hgAIN, and cancer).

Multiple imputations analyses were performed to test whether the single imputation led to any bias [52]. We used the multivariable imputation by chained equations (MICE) procedure (Classification and Regression Trees, CART) [53] with  $m = 100$  multiple imputations [54]. Rubin's rules were used to combine the multiply imputed estimates [55]. Receiver operating characteristics (ROC) of the DNAm score and AUC (95% CI) were estimated for the 100 multiple imputations separately and for their average.

All  $p$ -values were two-sided with significance set at  $\alpha < 0.05$ . Analyses were undertaken using R statistical software version 3.3.1 [56].

## Abbreviations

AIN: anal intraepithelial neoplasia; ASCUS: Atypical Squamous Cells of Undetermined Significance; AUC: area under the curve; CART: Classification and Regression Trees; CI: confidence interval; CIN: cervical intraepithelial neoplasia; CpG: 5'—Cytosine—phosphate—Guanine—3'; DNAm: DNA methylation; EDTA: Ethylenediaminetetraacetic acid; FFPE: formalin-fixed paraffin-embedded; HCl: Hydrochloric Acid; H&E: Hematoxylin and Eosin; hgAIN: high-grade anal intraepithelial neoplasia; hgCIN: high-grade cervical intraepithelial neoplasia; HIV: human immunodeficiency virus; HRA: High-resolution anoscopy; hrHPV: high-risk human papillomavirus; HSIL: high-grade squamous intraepithelial lesion; HPV: human papillomavirus; IQR: interquartile range; lgAIN: low-grade anal intraepithelial neoplasia; LR: likelihood ratio; lrHPV: low-risk human papillomavirus; LSIL: low-grade squamous intraepithelial lesion; MICE: multivariate imputation by chained equations; MLC: human cardiac myosin light chain; MSM: men who have sex with men; OR: odds ratio; PCR: Polymerase Chain Reaction; R&D: research & development; ROC: Receiver operating characteristics; SDS: Sodium Dodecyl Sulfate; Tris: trisaminomethane; UK: United Kingdom

## Authors' contributions

AL, MN, MT, JC and PS planned the study, supervised the statistical analyses and wrote the manuscript. MN and MT collected patient samples and data. CR designed the experimental work, performed the experimental work, analyzed the data and wrote the manuscript. RW performed the experimental work. MS histologically reviewed the biopsies. NV wrote the manuscript. AA analyzed the data and provided detailed statistical input. All authors read the paper and agree with its contents.

## ACKNOWLEDGMENTS

The authors gratefully recognize the patients who participated in this study. This work was supported by Cancer Research UK (grant number C569/A10404). The study sponsors had no role in the study design, in the collection, analysis and interpretation of data; in the writing of the manuscript; or in the decision to submit the manuscript for publication.

## CONFLICTS OF INTEREST

There are no conflicts of interest to disclose.

## FUNDING

Cancer Research UK programme grant (grant number C569/A10404).

## REFERENCES

1. Bosch FX, Lorincz A, Munoz N, Meijer CJLM, Shah KV. The causal relation between human papillomavirus and cervical cancer. *J Clin Pathol.* 2002; 55:244–65.
2. zur Hausen H. Papillomaviruses and cancer: from basic studies to clinical application. *Nat Rev Cancer.* 2002; 2:342–50. doi: 10.1038/nrc798.
3. Paavonen J, Naud P, Salmeron J, Wheeler CM, Chow SN, Apter D, Kitchener H, Castellsague X, Teixeira JC, Skinner SR, Hedrick J, Jaisamram U, Limson G, et al. Efficacy of human papillomavirus (HPV)-16/18 AS04-adjuvanted vaccine against cervical infection and precancer caused by oncogenic HPV types (PATRICIA): final analysis of a double-blind, randomised study in young women. *Lancet.* 2009; 374:301–14. doi: 10.1016/S0140-6736(09)61248-4.
4. Kjaer S, Hogdall E, Frederiksen K, Munk C, van den Brule A, Svare E, Meijer C, Lorincz A, Iftner T. The absolute risk of cervical abnormalities in high-risk human papillomavirus-positive, cytologically normal women over a 10-year period. *Cancer Res.* 2006; 66:10630–6. doi: 10.1158/0008-5472.CAN-06-1057.
5. de Pokomandy A, Rouleau D, Ghattas G, Vezina S, Cote P, Macleod J, Allaire G, Franco EL, Coutlee F, HIVIRG Study Group. Prevalence, Clearance, and Incidence of Anal Human Papillomavirus Infection in HIV-Infected Men: The HIVIRG Cohort Study. *J Infect Dis.* 2009; 199:965–73. doi: 10.1086/597207.
6. Cullen AP, Reid R, Campion M, Lorincz AT. Analysis of the Physical State of Different Human Papillomavirus DNAs in Intraepithelial and Invasive Cervical Neoplasm. *J Virol.* 1991; 65:606–12.
7. Bryant D, Onions T, Raybould R, Jones S, Tristram A, Hibbitts S, Fiander A, Powell N. Increased methylation of Human Papillomavirus type 16 DNA correlates with viral integration in Vulval Intraepithelial Neoplasia. *J Clin Virol.* 2014; 61:393–9. doi: 10.1016/j.jcv.2014.08.006.
8. Cuzick J, Bergeron C, Doeberitz MV, Gravitt P, Jeronimo J, Lorincz AT, Meijer CJLM, Sankaranarayanan R, Snijders PJF, Szarewski A. New Technologies and Procedures for Cervical Cancer Screening. *Vaccine.* 2012; 30: F107-F16. doi: 10.1016/j.vaccine.2012.05.088.
9. Elfstrom KM, Smelov V, Johansson ALV, Eklund C, Naucler P, Arnheim-Dahlstrom L, Dillner J. Long term duration of protective effect for HPV negative women: follow-up of primary HPV screening randomised controlled trial. *Brit Med J.* 2014; 348: g130. doi: ARTN g130 10.1136/bmj.g130.
10. Machalek DA, Poynten M, Jin FY, Fairley CK, Farnsworth A, Garland SM, Hillman RJ, Petoumenos K, Roberts J, Tabrizi SN, Templeton DJ, Grulich AE. Anal human papillomavirus infection and associated neoplastic lesions in men who have sex with men: a systematic review and meta-analysis. *Lancet Oncol.* 2012; 13:487–500. doi: 10.1016/S1470-2045(12)70080-3.

11. Silverberg MJ, Lau B, Justice AC, Engels E, Gill MJ, Goedert JJ, Kirk GD, D'Souza G, Bosch RJ, Brooks JT, Napravnik S, Hessel NA, Jacobson LP, et al. Risk of Anal Cancer in HIV-Infected and HIV-Uninfected Individuals in North America. *Clin Infect Dis*. 2012; 54:1026–34. doi: 10.1093/cid/cir1012.
12. Cancer Research UK. (2014). [http://www.cancerresearchuk.org/sites/default/files/cstream-node/cases\\_rates\\_anus\\_I14.pdf](http://www.cancerresearchuk.org/sites/default/files/cstream-node/cases_rates_anus_I14.pdf). Accessed January 2017.
13. Alemany L, Saunier M, Alvarado-Cabrero I, Quiros B, Salmeron J, Shin HR, Pirog EC, Guimera N, Hernandez-Suarez G, Felix A, Clavero O, Lloveras B, Kasamatsu E, et al. Human papillomavirus DNA prevalence and type distribution in anal carcinomas worldwide. *Int J Cancer*. 2015; 136:98–107. doi: 10.1002/ijc.28963.
14. Salit IE, Timmouth J, Chong S, Raboud J, Diong C, Su DS, Sano M, Lytwyn A, Chapman W, Mahony J. Screening for HIV-Associated Anal Cancer: Correlation of HPV Genotypes, p16, and E6 Transcripts with Anal Pathology. *Cancer Epidem Biomar*. 2009; 18:1986–92. doi: 10.1158/1055-9965.Epi-08-1141.
15. Palefsky JM, Holly EA, Hogeboom CJ, Berry JM, Jay N, Darragh TM. Anal cytology as a screening tool for anal squamous intraepithelial lesions. *J Acq Immun Def Syndr*. 1997; 14:415–22. doi:
16. Nathan M, Singh N, Garrett N, Hickey N, Prevost T, Sheaff M. Performance of anal cytology in a clinical setting when measured against histology and high-resolution anoscopy findings. *Aids*. 2010; 24:373–9. doi: 10.1097/QAD.0b013e328333ab8e.
17. Salit IE, Lytwyn A, Raboud J, Sano M, Chong S, Diong C, Chapman W, Mahony JB, Timmouth J. The role of cytology (Pap tests) and human papillomavirus testing in anal cancer screening. *Aids*. 2010; 24:1307–13. doi: 10.1097/QAD.0b013e3283339e592.
18. Wentzensen N, Follansbee S, Borgonovo S, Tokugawa D, Schwartz L, Lorey TS, Sahasrabudde VV, Lamere B, Gage JC, Fetterman B, Darragh TM, Castle PE. Human papillomavirus genotyping, human papillomavirus mRNA expression, and p16/Ki-67 cytology to detect anal cancer precursors in HIV-infected MSM. *Aids*. 2012; 26:2185–92. doi: 10.1097/QAD.0b013e328359f255.
19. Lorincz AT, Brentnall AR, Scibior-Bentkowska D, Reuter C, Banwait R, Cadman L, Austin J, Cuzick J, Vasiljevic N. Validation of a DNA methylation HPV triage classifier in a screening sample. *Int J Cancer*. 2016; 138:2745–51. doi: 10.1002/ijc.30008.
20. Lorincz AT. Virtues and Weaknesses of DNA Methylation as a Test for Cervical Cancer Prevention. *Acta Cytol*. 2016; 60:501–12. doi: 10.1159/000450595.
21. De Strooper LMA, Verhoef VMJ, Berkhof J, Hesselink AT, de Bruin HME, van Kemenade FJ, Bosgraaf RP, Bekkers RLM, Massuger LFAG, Melchers WJG, Steenbergen RDM, Snijders PJF, Meijer CJLM, et al. Validation of the FAM19A4/mir124-2 DNA methylation test for both lavage- and brush-based self-samples to detect cervical (pre)cancer in HPV-positive women. *Gynecol Oncol*. 2016; 141:341–7. doi: 10.1016/j.ygyno.2016.02.012.
22. Hesselink AT, Heideman DA, Steenbergen RD, Coupe VM, Overmeer RM, Rijkaart D, Berkhof J, Meijer CJ, Snijders PJ. Combined promoter methylation analysis of CADM1 and MAL: an objective triage tool for high-risk human papillomavirus DNA-positive women. *Clin Cancer Res*. 2011; 17:2459–65. doi: 10.1158/1078-0432.CCR-10-2548.
23. Lorincz AT, Brentnall AR, Vasiljevic N, Scibior-Bentkowska D, Castanon A, Fiander A, Powell N, Tristram A, Cuzick J, Sasieni P. HPV16 L1 and L2 DNA methylation predicts high-grade cervical intraepithelial neoplasia in women with mildly abnormal cervical cytology. *Int J Cancer*. 2013. doi: 10.1002/ijc.28050.
24. Mirabello L, Schiffman M, Ghosh A, Rodriguez AC, Vasiljevic N, Wentzensen N, Herrero R, Hildesheim A, Wacholder S, Scibior-Bentkowska D, Burk RD, Lorincz AT. Elevated methylation of HPV16 DNA is associated with the development of high grade cervical intraepithelial neoplasia. *Int J Cancer*. 2013; 132:1412–22. doi: 10.1002/ijc.27750.
25. Zhang Y, Xu R, Li G, Xie X, Long J, Wang H. Loss of expression of the differentially expressed in adenocarcinoma of the lung (DAL-1) protein is associated with metastasis of non-small cell lung carcinoma cells. *Tumour Biol*. 2012; 33:1915–25. doi: 10.1007/s13277-012-0452-x.
26. Heller G, Geradts J, Ziegler B, Newsham I, Filipits M, Markis-Ritzinger EM, Kandioler D, Berger W, Stiglbauer W, Depisch D, Pirker R, Zielinski CC, Zochbauer-Muller S. Downregulation of TSLC1 and DAL-1 expression occurs frequently in breast cancer. *Breast Cancer Res Treat*. 2007; 103:283–91. doi: 10.1007/s10549-006-9377-7.
27. Dafou D, Grun B, Sinclair J, Lawrenson K, Benjamin EC, Hogdall E, Kruger-Kjaer S, Christensen L, Sowter HM, Al-Attar A, Edmondson R, Darby S, Berchuck A, et al. Microcell-mediated chromosome transfer identifies EPB41L3 as a functional suppressor of epithelial ovarian cancers. *Neoplasia*. 2010; 12:579–89. doi:
28. Vasiljevic N, Scibior-Bentkowska D, Brentnall AR, Cuzick J, Lorincz AT. Credentialing of DNA methylation assays for human genes as diagnostic biomarkers of cervical intraepithelial neoplasia in high-risk HPV positive women. *Gynecol Oncol*. 2014; 132:709–14. doi: 10.1016/j.ygyno.2014.02.001.
29. Brentnall AR, Vasiljevic N, Scibior-Bentkowska D, Cadman L, Austin J, Cuzick J, Lorincz AT. HPV33 DNA methylation measurement improves cervical pre-cancer risk estimation of an HPV16, HPV18, HPV31 and EPB41L3 methylation classifier. *Cancer Biomark*. 2015; 15:669–75. doi: 10.3233/CBM-150507.
30. Chiao EY, Giordano TP, Palefsky JM, Tyring S, El Serag H. Screening HIV-infected individuals for anal cancer precursor lesions: A systematic review. *Clin Infect Dis*. 2006; 43:223–33. doi: 10.1086/505219.

31. Mathews WC, Sitapati A, Caperna JC, Barber RE, Tugend A, Go U. Measurement characteristics of anal cytology, histopathology, and high-resolution anoscopic visual impression in an anal dysplasia screening program. *J Acq Imm Def.* 2004; 37:1610–5. doi: 10.1097/00126334-200412150-00014.
32. Lytwyn A, Salit IE, Raboud J, Chapman W, Darragh T, Winkler B, Tinmouth J, Mahony JB, Sano M. Interobserver agreement in the interpretation of anal intraepithelial neoplasia. *Cancer.* 2005; 103:1447–56. doi: 10.1002/cncr.20927.
33. Goldstone SE, Lowe B, Rothmann T, Nazarenko I. Evaluation of the hybrid capture 2 assay for detecting anal high-grade dysplasia. *Int J Cancer.* 2012; 131:1641–8. doi: 10.1002/ijc.27431.
34. Giuliano AR, Lee JH, Fulp W, Villa LL, Lazcano E, Papenfuss MR, Abrahamsen M, Salmeron J, Anic GM, Rollison DE, Smith D. Incidence and clearance of genital human papillomavirus infection in men (HIM): a cohort study. *Lancet.* 2011; 377:932–40. doi: 10.1016/S0140-6736(10)62342-2.
35. Mooij SH, van Santen DK, Geskus RB, van der Sande MAB, Coutinho RA, Stolte IG, Snijders PJF, Meijer CJLM, Speksnijder AGCL, de Vries HJC, King AJ, van Eeden A, van der Loeff MFS. The effect of HIV infection on anal and penile human papillomavirus incidence and clearance: a cohort study among MSM. *Aids.* 2016; 30:121–32. doi: 10.1097/Qad.0000000000000909.
36. Heard I, Poizot-Martin I, Potard V, Etienney I, Crenn-Hebert C, Moore C, Touraine P, Cubie H, Costagliola D, ANRS-C017 VIHGY Study Group. Prevalence of and Risk Factors for Anal Oncogenic Human Papillomavirus Infection Among HIV-Infected Women in France in the Combination Antiretroviral Therapy Era. *J infect Dis.* 2016; 213:1455–61. doi: 10.1093/infdis/jiv751.
37. Serrano B, de Sanjose S, Tous S, Quiros B, Munoz N, Bosch X, Alemany L. Human papillomavirus genotype attribution for HPVs 6, 11, 16, 18, 31, 33, 45, 52 and 58 in female anogenital lesions. *Eur J Cancer.* 2015; 51:1732–41. doi: 10.1016/j.ejca.2015.06.001.
38. Louvanto K, Franco EL, Ramanakumar AV, Vasiljevic N, Scibior-Bentkowska D, Koushik A, Cuzick J, Coutlee F, Lorincz AT, Biomarkers of Cervical Cancer Risk Study Team. Methylation of viral and host genes and severity of cervical lesions associated with human papillomavirus type 16. *Int J Cancer.* 2015; 136:E638–45. doi: 10.1002/ijc.29196.
39. De Strooper LM, van Zummeren M, Steenbergen RD, Bleeker MC, Hesselink AT, Wisman GB, Snijders PJ, Heideman DA, Meijer CJ. CADM1, MAL and miR124-2 methylation analysis in cervical scrapes to detect cervical and endometrial cancer. *J Clin Pathol.* 2014; 67:1067–71. doi: jclinpath-2014-202616 [pii] 10.1136/jclinpath-2014-202616.
40. Bierkens M, Hesselink AT, Meijer CJ, Heideman DA, Wisman GB, van der Zee AG, Snijders PJ, Steenbergen RD. CADM1 and MAL promoter methylation levels in hrHPV-positive cervical scrapes increase proportional to degree and duration of underlying cervical disease. *Int J Cancer.* 2013; 133:1293–9. doi: 10.1002/ijc.28138.
41. Darragh TM, Colgan TJ, Cox JT, Heller DS, Henry MR, Luff RD, McCalmont T, Nayar R, Palefsky JM, Stoler MH, Wilkinson EJ, Zaino RJ, Wilbur DC, et al. The Lower Anogenital Squamous Terminology Standardization Project for HPV-associated Lesions: Background and Consensus Recommendations From the College of American Pathologists and the American Society for Colposcopy and Cervical Pathology. *Int J Gynecol Pathol.* 2013; 32:76–115. doi: 10.1097/PGP.0b013e31826916c7.
42. Klaes R, Friedrich T, Spitkovsky D, Ridder R, Rudy W, Petry U, Dallenbach-Hellweg G, Schmidt D, Doeberitz MV. Overexpression of p16(ink4a) as a specific marker for dysplastic and neoplastic epithelial cells of the cervix uteri. *Int J Cancer.* 2001; 92:276–84. doi: 10.1002/ijc.1174.
43. Vasiljevic N, Ahmad AS, Beesley C, Thorat MA, Fisher G, Berney DM, Moller H, Yu Y, Lu YJ, Cuzick J, Foster CS, Lorincz AT. Association between DNA methylation of HSPB1 and death in low Gleason score prostate cancer. *Prostate Cancer Prostatic Dis.* 2013; 16:35–40. doi: 10.1038/pcan.2012.47.
44. Goldenberger D, Perschil I, Ritzler M, Altwegg M. A simple “universal” DNA extraction procedure using SDS and proteinase K is compatible with direct PCR amplification. *PCR Methods Appl.* 1995; 4:368–70. doi:
45. Clifford GM, Rana RK, Franceschi S, Smith JS, Gough G, Pimenta JM. Human papillomavirus genotype distribution in low-grade cervical lesions: comparison by geographic region and with cervical cancer. *Cancer Epidemiol Biomarkers Prev.* 2005; 14:1157–64. doi: 14/5/1157 [pii] 10.1158/1055-9965.EPI-04-0812.
46. Hoots BE, Palefsky JM, Pimenta JM, Smith JS. Human papillomavirus type distribution in anal cancer and anal intraepithelial lesions. *Int J Cancer.* 2009; 124:2375–83. doi: 10.1002/ijc.24215.
47. Vasiljevic N, Scibior-Bentkowska D, Brentnall A, Cuzick J, Lorincz A. A comparison of methylation levels in HPV18, HPV31 and HPV33 genomes reveals similar associations with cervical precancers. *J Clin Virol.* 2014; 59:161–6. doi: 10.1016/j.jcv.2013.12.014.
48. Vasiljevic N, Wu K, Brentnall AR, Kim DC, Thorat MA, Kudahetti SC, Mao X, Xue L, Yu Y, Shaw GL, Beltran L, Lu YJ, Berney DM, et al. Absolute quantitation of DNA methylation of 28 candidate genes in prostate cancer using pyrosequencing. *Dis Markers.* 2011; 30:151–61. doi: 56GG82427WK228U8 [pii] 10.3233/DMA-2011-0790.
49. Janssen KJM, Donders ART, Harrell FE, Vergouwe Y, Chen QX, Grobbee DE, Moons KGM. Missing covariate data in medical research: To impute is better than to ignore. *J Clin Epidemiol.* 2010; 63:721–7. doi: 10.1016/j.jclinepi.2009.12.008.

50. DeLong ER, DeLong DM, Clarkepearson DI. Comparing the Areas under 2 or More Correlated Receiver Operating Characteristic Curves - a Nonparametric Approach. *Biometrics*. 1988; 44:837–45. doi: 10.2307/2531595.
51. Carpenter J, Bithell J. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Statistics in Medicine*. 2000; 19:1141–64. doi: 10.1002/(Sici)1097-0258(20000515)19:9<1141::Aid-Sim479>3.0.Co;2-F.
52. Rubin D, Little RJA. *Statistical Analysis with Missing Data*. (Hoboken, New Jersey: John Wiley & Sons).
53. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *J Stat Softw*. 2011; 45:1–67.
54. Hardt J, Herke M, Leonhart R. Auxiliary variables in multiple imputation in regression with missing X: a warning against including too many in small sample research. *Bmc Med Res Methodol*. 2012; 12. doi: Artn 184 10.1186/1471-2288-12-184.
55. Rubin DB. *Multiple imputation for nonresponse in surveys*. (Hoboken, N.J. ; Wiley-Interscience). 2004.
56. R Core Team. *R: A Language and Environment for Statistical Computing*. (Vienna, Austria: R Foundation for Statistical Computing, retrieved from <https://www.R-project.org>). 2016.